ON DISTRIBUTION-FREE STATISTICS

Astia

By

Z. W. Birnbaum and H. Rubin

University of Washington and Stanford University

Technical Report No. 11

July 24, 1953

Contract N8onr-520 Task Order II
Project Number NR-042-038

~~Laboratory of Statistical~~ Research
~~Department of Mathematics~~
University of Washington
Seattle, Washington

On Distribution-Free Statistics

by

Z. W. Birnbaum and H. Rubin [1]

University of Washington and Stanford University

1. Introduction.

Let $X_1, X_2, \ldots, X_n$ be a sample of a one-dimensional random variable X which has the continuous cumulative probability function F. It has been observed [1] that, to the authors' knowledge, all distribution-free statistics considered in the past can be written in the form $\Phi[F(X_1), F(X_2), \ldots, F(X_n)]$ where Φ is a measurable symmetric function defined on the unit-cube $\{U: 0 \leq U_i \leq 1,\ i = 1, 2, \ldots, n\}$. It is the purpose of this paper to study the relationship between the class of statistics which can be written in this particular form and the class of distribution-free statistics.

2. Distribution-free statistics and statistics of structure (d).

Let Ω and Ω' be two families of cumulative probability functions. A real quantity

$$W = S(X_1, X_2, \ldots, X_n, G)$$

will be called a <u>statistic in Ω with regard to Ω'</u> if, for any $G \in \Omega$, $F \in \Omega'$, and $X_1, X_2, \ldots, X_n$ in the n-dimensional sample-space for a random variable X which has the cumulative probability function F,

1° $S(X_1, X_2, \ldots, X_n, G)$ is defined almost everywhere in the sample-space $X_1, X_2, \ldots, X_n$ (i.e. with the possible exception of a set of probability zero), and

[1] Work done under the sponsorship of the Office of Naval Research.

2º $W = S(X_1, X_2, \ldots, X_n, G)$ has a probability distribution; this probability distribution will be denoted by

$\mathcal{P}(W;F) = \mathcal{P}\left[S(X_1, X_2, \ldots, X_n, G); F\right]$.

For example, Kolmogorov's statistic

$$D_n = \sup_{-\infty < x < \infty} \left| F_n(x) - G(x) \right|, \tag{2.1}$$

where F_n is the empirical cumulative distribution function determined by the sample $X_1, X_2, \ldots, X_n$, satisfies 1º and 2º when $\Omega = \Omega' = \Omega_1$, the class of all non-degenerate cumulative probability functions [2/], hence D_n is a statistic in Ω_1 with regard to Ω_1.

If for a statistic $S(X_1, X_2, \ldots, X_n, G)$ in Ω with regard to Ω' there exists a function Φ defined on the n-dimensional unit cube and symmetric in its arguments, such that for any $G \in \Omega$, $F \in \Omega'$ we have

$$S(X_1, X_2, \ldots, X_n, G) = \Phi\left[G(X_1), G(X_2), \ldots, G(X_n)\right]$$

almost everywhere [3/] in the sample space $X_1, X_2, \ldots, X_n$ for the random variable X which has the cumulative probability function F, then we shall say that $S(X_1, X_2, \ldots, X_n, G)$ is <u>a statistic of structure (d)</u>.

Kolmogorov's statistic (2.1) is an example of a statistic of structure (d), since it can be written as

$$D_n = \max_{i=1,\ldots,n} \left\{ \operatorname{Max}\left[G(X_i') - \frac{i-1}{n}, \frac{i}{n} - G(X_i')\right] \right\},$$

where $X_1', X_2', \ldots, X_n'$ are the numbers $X_1, X_2, \ldots, X_n$, ordered increasingly.

2/ The notations for various classes of cumulative probability functions are those introduced by Scheffé [2].

3/ The exceptional set of probability zero may depend on G.

If $\Omega = \Omega'$ and the statistic $S(X_1, X_2, \ldots, X_n, G)$ has the property that the probability distribution $\mathcal{P}[S(X_1, X_2, \ldots, X_n, G); G]$ is independent of G for $G \in \Omega$, we shall say that $S(X_1, X_2, \ldots, X_n, G)$ is a <u>distribution-free statistic in Ω</u>.

Let us now assume $\Omega = \Omega' = \Omega_2$, the class of all continuous cumulative probability functions. Denoting by R the rectangular distribution in (0,1) we have

$$\mathcal{P}\left\{\Phi\left[G(X_1), \ldots, G(X_n)\right]; G\right\} = \mathcal{P}\left\{\Phi(U_1, \ldots, U_n); R\right\}.$$

It follows that <u>if a statistic in Ω_2 with regard to Ω_2 has structure (d) then it is distribution-free in Ω_2</u>.

All distribution-free statistics considered in literature happen to have structure (d), with $\Omega = \Omega' = \Omega_2$. Nevertheless the conjecture that every distribution-free statistic, symmetric in $X_1, X_2, \ldots, X_n$, with $\Omega = \Omega' = \Omega_2$, must have structure (d) is not true. This can be seen from the following counter-example:

Let ω_1 and ω_2 be non-empty, mutually exclusive subsets of Ω_2 such that $\omega_1 \cup \omega_2 = \Omega_2$. Denoting by F_n again the empirical cumulative distribution function determined by a sample of size n, we define

$$S = \begin{cases} \sup\limits_{-\infty < x < \infty} \left[F(x) - F_n(x)\right] = S_1, & \text{if } F \in \omega_1 \\ \sup\limits_{-\infty < x < \infty} \left[F_n(x) - F(x)\right] = S_2, & \text{if } F \in \omega_2 . \end{cases}$$

Since S_1 and S_2 are distribution-free statistics with the same probability distribution, S is a distribution-free statistic. It is, however, clearly not a statistic of structure (d).

3. Strongly distribution-free statistics.

Let Ω^* be the family of all continuous cumulative probability functions such that if $G \in \Omega^*$ then G is strictly increasing at all x

for which $0 < G(x) < 1$. Clearly if $G \in \Omega^*$ then the inverse function $G^{(-1)}$ is defined on the open unit interval.

We now consider a statistic $S(X_1, X_2, \ldots, X_n, G)$ in Ω^* with regard to some family Ω' of cumulative probability functions. This statistic shall be called <u>strongly-distribution-free in Ω^* with regard to Ω'</u> if the probability distribution $\mathcal{P}\left[S(X_1, X_2, \ldots, X_n, G); F\right]$ depends only on the function $\tau = F\,G^{(-1)}$ for all $G \in \Omega^*$, $F \in \Omega'$.

It is easily seen that, for $\Omega' = \Omega^*$, <u>a strongly distribution-free statistic is distribution-free.</u> For if $\mathcal{P}\left[S(X_1, X_2, \ldots, X_n, G); F\right]$ depends only on $F\,G^{(-1)}$ for all F, $G \in \Omega^*$, then in particular $\mathcal{P}\left[S(X_1, X_2, \ldots, X_n, G); G\right]$ depends only on $G\,G^{(-1)} = I$, hence is independent of G. One also verifies immediately that <u>if a statistic in Ω^* with regard to Ω^* has structure (d) then it is strongly distribution-free</u>, since then $\mathcal{P}\left\{\Phi\left[G(X_1), G(X_2), \ldots, G(X_n)\right]; F\right\}$ $= \mathcal{P}\left\{\Phi\left[U_1, U_2, \ldots, U_n\right]; F\,G^{(-1)}\right\}$.

Since all practically important distribution-free statistics are symmetric in $X_1, X_2, \ldots, X_n$ and strongly distribution-free, as well as of structure (d), one again may conjecture that under some fairly general assumptions these two properties are equivalent. This conjecture is found to be correct for $\Omega = \Omega' = \Omega^*$. We have already seen that if a statistic has structure (d) it is strongly distribution-free; it remains only to prove the converse statement:

<u>Theorem. If a statistic $V = S(X_1, X_2, \ldots, X_n, G)$ in Ω^* with regard to Ω^* is symmetric in $X_1, X_2, \ldots, X_n$ and strongly distribution-free, then it has structure (d).</u>

The proof of this theorem makes use of a lemma which will be presented in the next section.

4. A lemma.

Let H be a strictly increasing continuous function on the closed unit-interval, such that $H(0) = 0$, $H(1) = 1$; μ_H the measure defined by H on the unit-interval I_1; $\mu_H^{(n)}$ the corresponding product-measure on the n-dimensional unit-cube I_n. Then, for any set $M \subset I_n$ with $\mu_H^{(n)}(M) > 0$ and any $\varepsilon > 0$, there exist sets $Q_1, Q_2, \ldots, Q_n$ in I_1 such that

1º $Q_1, Q_2, \ldots, Q_n$ are disjoint, μ_H-measurable, and
$\mu_H(Q_i) > 0$, $\quad i = 1, 2, \ldots, n$,

2º for $Q_o = \text{Compl.} \bigcup\limits_{i=1} Q_i$ we have $\mu_H(Q_o) > 0$,

3º if Q_i is placed on the y_i-axis, $i = 1, 2, \ldots, n$, then the product-set $Q = Q_1 \times Q_2 \times \ldots \times Q_n$ in I_n has the property

$$\frac{\mu_H^{(n)}(Q \cap M)}{\mu_H^{(n)}(Q)} > 1 - \varepsilon.$$

Proof: it may be assumed without loss of generality that $H(y) = y$, so that μ_H and $\mu_H^{(n)}$ are Lebesgue measures. Let $C_{\eta, y_1, \ldots, y_n}$ denote the cube $|Y_i - y_i| < \eta$ in the $(Y_1, Y_2, \ldots, Y_n)$ space, with the center $(y_1, y_2, \ldots, y_n)$, and the volume $\mu_H^{(n)}(C_{\eta, y_1, \ldots, y_n}) = (2\eta)^n$.

It is well known that

$$(4.1) \qquad \lim_{\eta \to 0} (2\eta)^{-n} \mu_H^{(n)}(M \cap C_{\eta, y_1, \ldots, y_n}) = 1$$

for almost all points in M (see e.g. [2] p. 129). The subset of those points of M for which no two coordinates are equal and none is 0 or 1 has the same measure as M. Let M_1 be the set of all points of M for which (4.1) holds and which have no two coordinates equal and no coordinate 0 or 1.

Then $\mu_H^{(n)}(M_1) = \mu_H^{(n)}(M) > 0$. Let $y_1^o, \ldots, y_n^o$ be a point in M_1, and let

$$\lambda = \min\left\{\min_{(i)} y_i^o, \min_{(i)} (1 - y_i^o), \min_{i \neq j} |y_i^o - y_j^o|\right\}.$$

Clearly $0 < \lambda < \frac{1}{2}$, and for $0 < \eta < \frac{\lambda}{2}$ the intervals

(4.2) $$Q_i: \ (y_i^o - \eta, y_i^o + \eta), \quad i = 1, 2, \ldots, n,$$

are all in I_1 and satisfy 1^o and 2^o. If Q_i is placed on the Y_i-axis then the product-set $Q = Q_1 X Q_2 X \ldots X Q_n$ is the cube $C_{\eta, y_1^o, \ldots, y_n^o}$.

According to (4.1) there exists an $\eta_o > 0$ such that

$$(2\eta)^{-n} \mu_H^{(n)}(M \cap C_{\eta, y_1^o, \ldots, y_n^o}) > 1 - \varepsilon$$

for $\eta < \eta_o$. Choosing $\eta < \min(\eta_o, \frac{\lambda}{2})$ and constructing the intervals (4.2) one obtains the Q_i required by the Lemma.

5. Proof of Theorem.

When the random variable X has the cumulative probability function F, the random variable $Y = G(X)$ has the cumulative probability function $H = F\, G^{(-1)}$. Setting $Y_i = G(X_i)$ we, therefore, have

$$W = S(X_1, \ldots, X_n, G) = S\left[G^{(-1)}(Y_1), \ldots, G^{(-1)}(Y_n), G\right]$$

and

$$\mathcal{P}\left[S(X_1, \ldots, X_n, G);\ F\right] = \mathcal{P}\left\{S\left[G^{(-1)}(Y_1), \ldots, G^{(-1)}(Y_n), G\right];\ FG^{(-1)}\right\} =$$
$$= \mathcal{P}\left\{S\left[G^{(-1)}(Y_1), \ldots, G^{(-1)}(Y_n), G\right];\ H\right\}.$$

By assumption, this last probability distribution depends only on the cumulative probability function H, and not on G. From this and the symmetry assumption we wish to conclude that $S\left[G^{(-1)}(Y_1), \ldots, G^{(-1)}(Y_n), G\right]$

can be written in the form of a function $\Phi(Y_1,\ldots,Y_n)$, independent of G except on a set of H-measure zero.

To prove this, we assume that for some $G_1, G_2 \in \Omega^*$ we have $S\left[G_1^{(-1)}(Y_1),\ldots,G_1^{(-1)}(Y_n),G_1\right] \neq S\left[G_2^{(-1)}(Y_1),\ldots,G_2^{(-1)}(Y_n),G_2\right]$ on a set of positive H-measure. Without loss of generality we may assume

(5.1) $\infty > k > S\left[G_1^{(-1)}(Y_1),\ldots,G_1^{(-1)}(Y_n),G_1\right] - S\left[G_2^{(-1)}(Y_1),\ldots,G_2^{(-1)}(Y_n),G_2\right] > \eta > 0$

on a set M in the unit cube I_n, where M is symmetric and has positive measure. For any H, continuous and strictly increasing in I_1, and any $\varepsilon > 0$, we construct sets $Q_1, Q_2, \ldots, Q_n$ according to the Lemma in Section 4 and have

(5.2)
$$\frac{\mu_H^{(n)}(Q\cap M)}{\mu_H^{(n)}(Q)} > 1 - \varepsilon.$$

For any

(5.3)
$$\alpha_i > 0, \qquad i = 0,1,\ldots,n$$
$$\alpha_0 + \sum_{i=1}^{n} \alpha_i = 1$$

we define the set function

$$K_{\alpha_1,\ldots,\alpha_n}(T) = \sum_{j=0}^{n} \alpha_j \frac{\mu_H(T\cap Q_j)}{\mu_H(Q_j)}$$

for any measurable $T \subset I_1$. This clearly is a probability measure in I_1. Taking for T the interval $(0,y)$ we obtain a strictly increasing continuous cumulative probability function which will be denoted by $K_{\alpha_1,\ldots,\alpha_n}$.

Without loss of generality, S may be assumed bounded, since otherwise

we could consider $\frac{S}{1+|S|}$. This assures the existence of the mathematical expectation of S. Since $S\left[G_1^{(-1)}(Y_1),\dots,G_1^{(-1)}(Y_n),G_1\right]$ and $S\left[G_2^{(-1)}(Y_1),\dots,G_2^{(-1)}(Y_n),G_2\right]$ have the same probability distribution if $Y_1,Y_2,\dots,Y_n$ are a sample of a random variable Y with the cumulative probability function $K_{\alpha_1,\dots,\alpha_n}$, their mathematical expectations are equal

$$(5.4)\quad E\left\{S\left[G_1^{(-1)}(Y_1),\dots,G_1^{(-1)}(Y_n),G_1\right]-S\left[G_2^{(-1)}(Y_1),\dots,G_2^{(-1)}(Y_n),G_2\right];K_{\alpha_1,\dots,\alpha_n}\right\}=0.$$

Using the abbreviations

$$S\left[G_i^{(-1)}(Y_1),\dots,G_i^{(-1)}(Y_n),G_i\right]=S_i(Y_1,\dots,Y_n),\qquad i=1,2,$$

we write the left-hand side of (5.4) explicitly

$$\int_{Y_1=0}^{1}\cdots\int_{Y_n=0}^{1}\left[S_1(Y_1,\dots,Y_n)-S_2(Y_1,\dots,Y_n)\right]\prod_{i=1}^{n} dK_{\alpha_1,\dots,\alpha_n}(Y_i)=$$

$$(5.5)=\sum_{j_1=0}^{n}\cdots\sum_{j_n=0}^{n}\int_{Y_1\in Q_{j_1}}\cdots\int_{Y_n\in Q_{j_n}}\left[S_1(Y_1,\dots,Y_n)-S_2(Y_1,\dots,Y_n)\right]\prod_{i=1}^{n} dK_{\alpha_1,\dots,\alpha_n}(Y_i)=$$

$$=\sum_{j_1=0}^{n}\cdots\sum_{j_n=0}^{n}\frac{\alpha_{j_1}\cdots\alpha_{j_n}}{\mu_H(Q_{j_1})\cdots\mu_H(Q_{j_n})}\int_{Q_{j_1}}\cdots\int_{Q_{j_n}}\left[S_1(Y_1,\dots,Y_n)-S_2(Y_1,\dots,Y_n)\right]$$

$$dH(Y_n)\dots dH(Y_1).$$

Since $S_1(Y_1,\dots,Y_n)$, $S_2(Y_1,\dots,Y_n)$ and M are symmetric in $Y_1,\dots,Y_n$, all the terms of the sum which correspond to different permutations of the same n subscripts $j_1,\dots,j_n$ (out of the $n+1$ possible values $0,1,\dots,n$) are equal. Collecting these equal terms, we obtain a polynomial in $\alpha_0,\alpha_1,\dots,\alpha_n$, which according to (5.4) vanishes identically under the

restrictions (5.3). It follows that each of the integrals in the last term of (5.5) must vanish, and in particular

$$\int_{Q_1}\int_{Q_2}\cdots\int_{Q_n}\left[S_1(Y_1,Y_2,\ldots,Y_n) - S_2(Y_1,Y_2,\ldots,Y_n)\right]dY_n\ldots dY_2dY_1 = 0;$$

which, for ε sufficiently small, contradicts (5.1) and (5.2).

References

1 Z. W. Birnbaum, "Distribution-free tests of fit for continuous distribution functions", Ann. Math. Stat., Vol. 24 (1953), pp. 1-7.

2 S. Saks, "Theory of the integral", Monografie Matematyczne, v. 7, Warszawa-Lwow, 1937.